Longitudinal Metabolomics data analysis using Weighted PCA

Jeroen J. Jansen[1], Hans F.M. Boelens[1], Huub C.J. Hoefsloot[1], Jan van der Greef[2,3], Age K. Smilde[1,2]

1) Process Analysis and Chemometrics, Faculty of Sciences, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands

2) TNO Nutrition and Food Reseach, PO Box 360, 3700 AJ Zeist, The Netherlands

3) Beyond Genomics, 40 Bear Hill Road, Waltham, MA 02451, United States

Website: http://www.science.uva.nl/its/ E-Mail: jjansen@science.uva.nl

Introduction

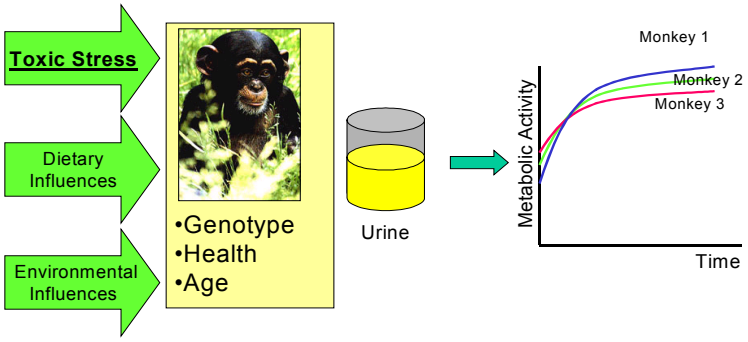
Metabolomics:

Research into the metabolism of an organism, influenced by **Internal** and **External** Factors.

How?: Analysis of chemical composition of Body Fluids (**urine**)

Using?: ^{1}H-NMR spectroscopy and multivariate data analysis methods

This Research : Analysing **normal** metabolism of Rhesus Monkeys



Methods

Methods give a **lower dimensional** representation of **high-dimensional** data.

Principal Component Analysis

Minimize:

$$f_{PCA} = \left\| \mathbf{X} - \mathbf{TP}' \right\|^2$$, (**T** are the scores, **P** are the loadings)

- Describes maximum variation in the data
- Additional Information can not always be used in PCA

Weighted Principal Component Analysis:

Minimize:

$$f_{WPCA} = \left\| \mathbf{W} * \left(\mathbf{X} - \mathbf{TP}' \right) \right\|^2$$

W : can be defined using:
- Missing Data
- Data Scaling
- **Experimental error**
- Other a priori information

- Does not describe maximum variation in the data
- Additional Information can be used by defining weights

Algorithms:

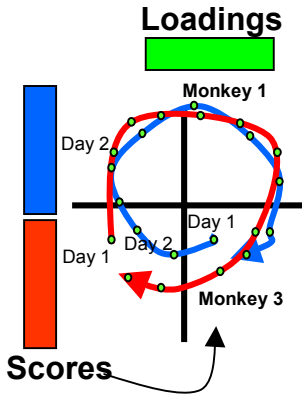
MLPCA	:	Wentzell, P. D. et al. (1997)
MILES	:	Bro, R. et al. (2002)
PCAW	:	Kiers, H. A. L. (1997)

Model

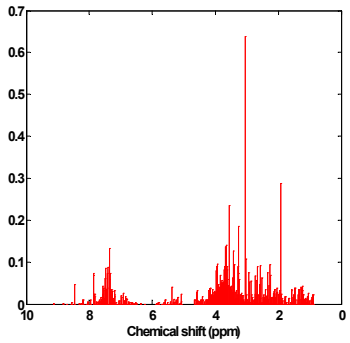
Model: normal *temporal* variation in the chemical composition of the **urine**.

Loadings: space that is spanned by the **typical** composition of the **urine** of **all monkeys**.

Scores: each urine **sample** in terms of space spanned by loadings. One score for each urine at each sampling time.



Dataset



Urine of **10 monkeys**
- 5 male
- 5 female.

Sampled 29 days in 2 months.

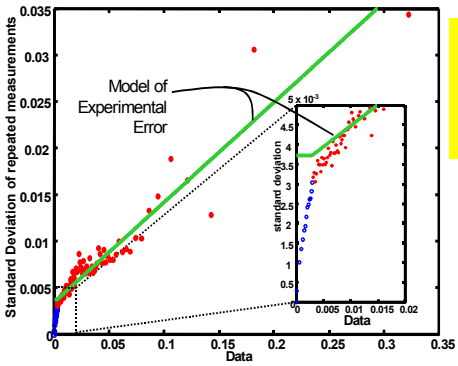
3 repeated measurements of each sample

Measurements:
NMR-spectra containing 332 chemical shifts

Repeated measurements:

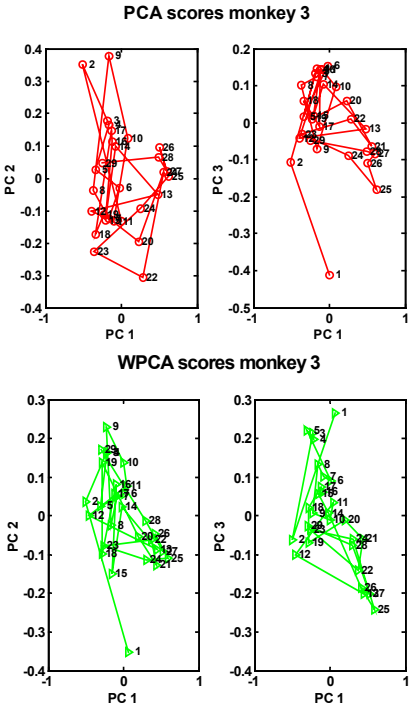
Additional information about the data

Standard Deviation = Estimation of **Experimental Error**



W :
Reciprocal value of the model of the experimental error

Results



Results of WPCA and PCA are similar

There are differences between the PCA and WPCA scores

The weighing causes the difference between the scores

The weighing that is defined here is indirectly based on peak size

Smaller peaks are given a larger weight, which gives them more importance in the data analysis

PCA and WPCA give a different view on the data, both methods can be combined to obtain a more complete view of the underlying processes

References:

Wentzell, P. D. Et al. (1997) Maximum likelihood principal component analysis. J. Chemom. 11, 339-366

Bro, R. Et al. (2002) Maximum Likelihood fitting using ordinary least squares algorithms. J. Chemom. 16, 387-400

Kiers, H. A. L. (1997) Weighted least squares fitting using ordinary least squares algorithms. Psychometrika, 62, 251-266